

Journal Pre-proofs

Research paper

Genetic Variations and Haplotypic Diversity in the Myostatin Gene of New Zealand Cattle Breeds

Ishaku L. Haruna, Ugonna J. Ekegbu, Farman Ullah, Hamed Amirpour-Najafabadi, Huitong Zhou, Jon G.H. Hickford

PII: S0378-1119(20)30069-X
DOI: <https://doi.org/10.1016/j.gene.2020.144400>
Reference: GENE 144400

To appear in: *Gene Gene*

Received Date: 9 January 2019
Revised Date: 17 January 2020
Accepted Date: 23 January 2020



Please cite this article as: I.L. Haruna, U.J. Ekegbu, F. Ullah, H. Amirpour-Najafabadi, H. Zhou, J.G.H. Hickford, Genetic Variations and Haplotypic Diversity in the Myostatin Gene of New Zealand Cattle Breeds, *Gene Gene* (2020), doi: <https://doi.org/10.1016/j.gene.2020.144400>

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2020 Published by Elsevier B.V.

ABREVIATION LIST

NZ	New Zealand
PCR	Polymerase Chain Reaction
SSCP	Single Strand Conformational Polymorphism
<i>MSTN</i>	The myostatin gene
MSTN	Myostatin protein
GDF8	Growth and Differentiation Factor 8
SNP	Single Nucleotide Polymorphism

ABSTRACT

Myostatin (MSTN) is a circulating factor that is secreted by muscle cells, and that acts upon those cells to inhibit the proliferation of muscle fibres during pre-natal muscle growth. The Polymerase Chain Reaction (PCR) coupled with Single Strand Conformational Polymorphism (SSCP) analysis, was used to reveal variation in the bovine MSTN gene (*MSTN*) in 722 cattle from a variety of breeds farmed in New Zealand (NZ). These included Hereford, Angus, Charolais, Simmental, Red Poll, South Devon, Shorthorn, Murray Grey, cross-bred Holstein-Friesian x Jersey cattle, and other composite breeds of cattle. Sequence analysis of five regions of *MSTN* that encompassed coding and non-coding regions of the gene, revealed a total of twelve single-nucleotide substitutions (7 in intron 1 and 5 in a region spanning the intron 2 - exon 3 boundary), and a single nucleotide deletion. Of these 12 substitutions, five are reported here for the first time, whereas seven have been previously described. The deletion c.748-78del, was located in the intron 2 - exon 3 boundary region, and has been reported previously. No nucleotide variation was identified in exons 1, 2 and 3. A total of 18 extended haplotypes were resolved spanning two variable regions (intron 1 and the intron 2 - exon 3 boundary), some of which were common across the breeds, while others were peculiar to particular breeds. The genetic variations identified provide insight into the conserved and polymorphic nature of the coding and non-coding sequences of bovine *MSTN* respectively, and thus provides a baseline for further study into how variation in the gene might affect growth and carcass traits in NZ cattle.

Keywords

Myostatin (MSTN); nucleotide sequence variation; haplotype; cattle; introns; exons

Author contributions

- Ishaku Lemu Haruna designed the primers, designed the research, performed experiments and wrote the paper.
- Ugonna Ekegbu assisted in optimizing experimental conditions, data collection and writing the paper.
- Hamed Amirpour-Najafabadi assisted in performing experiments and editing
- Farman Ullah assisted in performing the experiments and writing the paper
- Huitong Zhou and Jon Hickford both designed and supervised the research, assisted in interpreting results and editing the paper.

Conflict of interest

The authors declare no conflict of interest.

Declaration of interests

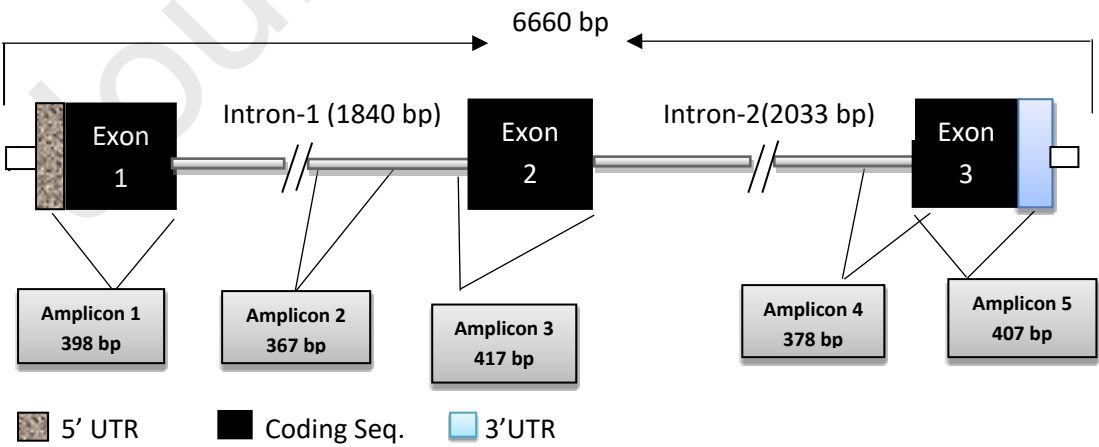
☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:



HIGHLIGHTS:

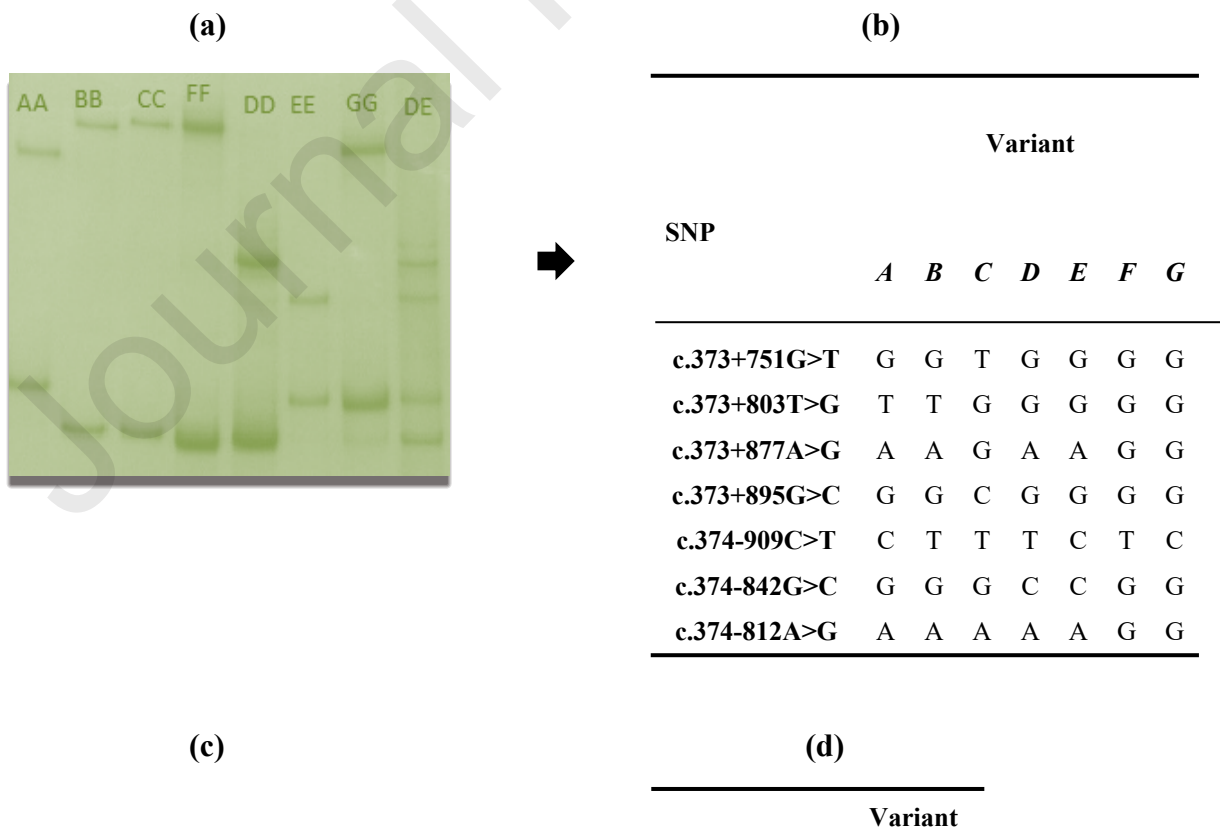
- The intronic regions of *MSTN* in New Zealand cattle breeds are highly polymorphic
- The exons of *MSTN* in New Zealand cattle breeds are conserved
- Linkage of bovine *MSTN* haplotypes based on proximal occurrence of SNP's
- Founder effect, origin and selection pressure could explain differences within cattle breeds.



Amplicon	Forward Primer	Annealing	Polyacrylamide	Optimized SSCP	Number of	GenBank
----------	----------------	-----------	----------------	----------------	-----------	---------

	Reverse Primer	Temperature (°C)	concentration (%)	Condition	SSCP Patterns	Accession number
1	F 5' ggaagaagtaagaacaaggg 3' R 5' caccagcaggactactcac 3'	60	12	280 V, 19 hr, 13 °C	1	
2	F 5' catgttactattgttgagag 3' R 5' aaggcaaatctattccagg 3'	58	14	390 V, 19 hr, 7 °C	7	MK353501-MK353507
3	F 5' gatatggaggtgtttgttcg 3' R 5' cagtaatcacttaccagtcc 3'	58	14	390 V, 19 hr, 13 °C	1	
4	F 5' ataagcaggaagacataagc 3' R 5' ggtgtgtctgttaccttgac 3'	58	14	390 V, 19 hr, 7 °C	3	MK353508-MK353510
5	F 5' ctcttcttcttccatcac 3' R 5' aagacctccatgtttgag 3'	60	11	200 V, 19 hr, 20 °C	1	

Figure 1. Primer pairs and amplicons used for the PCR-SSCP analysis of bovine *MSTN*. Primers were derived from bovine *MSTN* sequence with GenBank accession number AB076403. Optimized PCR-SSCP conditions for each amplicon and banding patterns identified are also shown.



BREED	SNP								c.748-78del	c.748-195C>T
	A B C									
	c.748-78del - - T									
	c.748-196C>T C C T									
	Intron 1									
	c.373+751#	c.373+803	c.373+877	c.373+895	c.374-909	c.374-842	c.374-812			
Hereford	G/T	T/G	A/G	G/C	C/T	G/C	A	-/T	C/T	
Angus	G	T/G	A	G	C/T	C/C	A	-/T	C/T	
South Devon	G	T/G	A	G	C/T	G/C	A	-/T	C/T	
Charolais	G/T	T/G	A/G	G/C	C/T	G/C	A	-/T	C/T	
Red Poll	G/T	T/G	A/G	G/C	C/T	G	A	-	C	
Shorthorn	G/T	T/G	A/G	G/C	C/T	G	A/G	-/T	C/T	
Simmental	G	T	A	G	C/T	G	A	-/T	C/T	
Composite	G/T	T/G	A/G	G/C	C/T	G	A	-	C	
Murray Grey	G/T	T/G	A/G	G/C	C/T	G/C	A	-/T	C/T	
Holstein-Friesian x Jersey cross-bred	G/T	T/G	A/G	G/C	C/T	G/C	A/G	-/T	C/T	

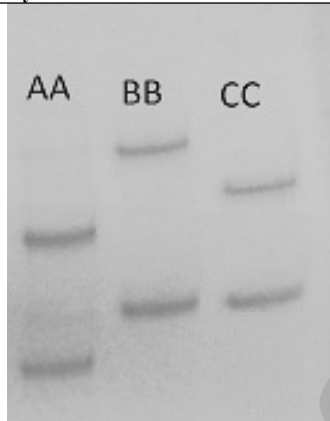


Figure 2. a) PCR-SSCP patterns for a 367 bp fragment of the intron 1 region of bovine *MSTN*. Seven different banding patterns (*A*, *B*, *C*, *D*, *E*, *F*, and *G*) were identified, with homozygous patterns for these variants being shown, along with a heterozygous (*DE*) pattern. b) Sequence analysis revealed that these variants were defined by seven different nucleotide variations. c) PCR-SSCP patterns for a 378 bp fragment of the intron 2-exon 3 region of bovine *MSTN*. Three banding patterns representing three variants (*A*, *B* and *C*) in homozygous form were identified. d) Sequence analysis revealed that these variants were defined by six nucleotide variations, including a deletion.

Table 1. Nucleotide variation within bovine *MSTN* in ten NZ cattle breeds and crosses

* The symbol (-) represents a deletion

Numbering is based on the recommended nucleotide nomenclature (<http://www.hgvs.org/mutnomen/recs-DNA.html#number>)

Table 2. The frequency of the *MSTN* haplotypes identified in the NZ cattle breeds

[§] Except for the Composite cattle, as a consequence of the absence of a cattle with a homozygous genotype that would enable haplotype determination.

[#] Symbol (-) signifies the absence of that haplotype within the breed.

BREED [§]	Frequency of each haplotype [#]												
	n	H1	H2	H3	H4	H5	H6	H7	H8	H9	H10	H11	H12
Hereford	28	7.14	21.43	-	-	3.57	7.14	17.86	3.57	-	21.43	3.57	-
Angus	9	-	50	6.25	-	12.5	18.8	-	-	-	-	-	-
South Devon	3	-	-	-	-	-	50	-	-	-	-	-	-
Charolaise	9	12.5	75	-	-	-	-	-	6.25	-	-	-	-
Red Poll	12	25	33.33	-	16.67	16.7	-	8.33	-	-	-	-	-
Shorthorn	17	-	86.35	-	-	-	-	-	4.55	4.55	-	-	-
Simmental	14	8.3	4.2	-	20.8	12.5	54.2	-	-	-	-	-	-
Murray Grey	17	4.55	72.7	-	-	-	-	9.1	9.1	4.55	-	-	-
Holstein-Friesian x Jersey cross-bred	36	9.2	11.11	3.7	11.11	9.2	5.5	18.47	7.5	-	1.9	-	3.7

Table 3. Nucleotide sequence variation identified in the 18 haplotypes (H1-H18) defined across the two regions of *MSTN* investigated

*The haplotypes can be clustered into seven broad groups based on the proximal occurrence of the sequence variation. In the intron 1 region, H1-H6 cluster as group 1; H7-H9 cluster as group 2; H10-H14 cluster as group 3; and H15-H18 cluster as group 4. In the intron 2/exon 3 region, H3, H6, H9, H12, H14, H15, and H18 cluster as group 5; H1, H4, H7, H10 and H16 cluster as group 6; and H2, H5, H8, H11, H13 and H17 cluster as group 7.

The symbol (-) represents a deletion and the highlighted nucleotides indicate the variation.

Haplo type*	Sequence Variation#												
	c.373 +751	c.373 +803	c.373 +877	c.373 +895	c.374- 909	c.374- 842	c.374- 812	c.748- 78del1	c.748- 195	c.748- 196	c.748- 281	c.748- 350	c.748- 352
H1	G	T	A	G	C	G	A	↓	C	C	G	C	T
H2	G	T	A	G	C	G	A	↓	C	C	C	C	C
H3	G	T	A	G	C	G	A	T	T	T	C	T	C
H4	G	T	A	G	T	G	A	↓	C	C	G	C	T
H5	G	T	A	G	T	G	A	↓	C	C	C	C	C
H6	G	T	A	G	T	G	A	T	T	T	C	T	C
H7	T	G	G	C	T	G	A	↓	C	C	G	C	T
H8	T	G	G	C	T	G	A	↓	C	C	C	C	C
H9	T	G	G	C	T	G	A	T	T	T	C	T	C
H10	G	G	A	G	T	C	A	↓	C	C	G	C	T
H11	G	G	A	G	T	C	A	↓	C	C	C	C	C
H12	G	G	A	G	T	C	A	T	T	T	C	T	C
H13	G	G	A	G	C	C	A	↓	C	C	C	C	C
H14	G	G	A	G	C	C	A	T	T	T	C	T	C
H15	G	G	G	G	T	G	G	T	T	T	C	T	C
H16	G	G	G	G	C	G	G	↓	C	C	G	C	T
H17	G	G	G	G	C	G	G	↓	C	C	C	C	C
H18	G	G	G	G	C	G	G	T	T	T	C	T	C

Genetic Variations and Haplotypic Diversity in the Myostatin Gene of New Zealand Cattle Breeds.

Ishaku L. Haruna*, Ugonna J. Ekegbu, Farman Ullah[#], Hamed Amirpour-Najafabadi, Huitong Zhou and Jon G. H. Hickford

Faculty of Agriculture and Life Sciences, Lincoln University, Lincoln 7647, New Zealand

[#]Department of Biotechnology, University of Malakand, Chakdara, Dir lower 18800, Pakistan

*Corresponding author(s): ishaku.haruna@Lincolnuni.ac.nz; Jonathan.Hickford@Lincoln.ac.nz

Tel. +64-27535-4073; +64-3-423-0665 (J.G.H.H.)

Abstract

Myostatin (MSTN) is a circulating factor that is secreted by muscle cells, and that acts upon those cells to inhibit the proliferation of muscle fibres during pre-natal muscle growth. The Polymerase Chain Reaction (PCR) coupled with Single Strand Conformational Polymorphism (SSCP) analysis, was used to reveal variation in the bovine MSTN gene (*MSTN*) in 722 cattle from a variety of breeds farmed in New Zealand (NZ). These included Hereford, Angus, Charolais, Simmental, Red Poll, South Devon, Shorthorn, Murray Grey, cross-bred Holstein-Friesian x Jersey cattle, and other composite breeds of cattle. Sequence analysis of five regions of *MSTN* that encompassed coding and non-coding regions of the gene, revealed a total of twelve single-nucleotide substitutions (7 in intron 1 and 5 in a region spanning the intron 2 - exon 3 boundary), and a single nucleotide deletion. Of these 12 substitutions, five are reported here for the first time, whereas seven have been previously described. The deletion c.748-78del, was located in the intron 2 - exon 3 boundary region, and has been reported previously. No nucleotide variation was identified in exons 1, 2 and 3. A total of 18 extended haplotypes were resolved spanning two variable regions (intron 1 and the intron 2 - exon 3 boundary), some of which were common across the breeds, while others were peculiar to particular breeds. The genetic variations identified provide insight into the

conserved and polymorphic nature of the coding and non-coding sequences of bovine *MSTN* respectively, and thus provides a baseline for further study into how variation in the gene might affect growth and carcass traits in NZ cattle.

Keywords

Myostatin (MSTN); nucleotide sequence variation; haplotype; cattle; introns; exons

1. Introduction

With an increasing human population comes the need for improved agricultural technologies to cater for the growing global demand for food. In this context, the meat industry would benefit from the development of improved ways of producing meat with increased efficiency. Having a good understanding of the genes that underpin animal growth and carcass traits is therefore of paramount importance.

One gene of recognised importance to animal growth and carcass traits is the myostatin (MSTN) gene (*MSTN*), sometimes called the Growth and Differentiation Factor 8 (GDF8) gene (*GDF8*). Myostatin is a circulating factor, secreted by muscle cells, and whose function is to regulate the pre-natal proliferation of muscle fibres (McPherron *et al.* 1997; Gonzalez-Cadavid *et al.* 1998). Genetic variation in *MSTN* has been identified in cattle and linked to having increased numbers of muscle fibres, or so-called ‘double-muscling’, in a number of breeds (Kambadur *et al.* 1997; McPherron and Lee, 1997, Grobet *et al.* 1997; Grisolia *et al.* 2009). At least 20 different genetic variants (deletions, insertions and nucleotide substitutions) have been described in cattle *MSTN* (Aiello *et al.* 2018).

While several studies have provided evidence of genetic variation in bovine *MSTN* and its effect on growth and carcass traits, most of these studies have been focused on investigating specific gene regions in European breeds. For example the seminal work of McPherron and Lee (1997) and Kambadur *et al.* (1997), described *MSTN* variation in Belgian Blue and Piedmontese Cattle respectively. The Belgian Blue cattle *MSTN* had an 11-nucleotide deletion in the third exon that caused a frameshift mutation, and the Piedmontese cattle *MSTN* had a missense mutation in exon 3, which was also associated with increased muscle. In their study, Kambadur *et al.* (1997), reported that upon finding the 11-nucleotide deletion, further pedigrees of Belgian Blue cattle were tested for this deletion, including 16 pedigrees

in New Zealand (NZ) and four in the United States of America (USA). Their results revealed that all the double-muscled purebred animals tested, were homozygous for this deletion.

Grobet *et al.* (1998) analysed ten European beef cattle breeds in which double-muscled animals have been described: including Belgian Blue cattle from Belgium; Blonde d'Aquitaine, Charolais, Gasconne, Limousin, Maine-Anjou, and Parthenaise cattle from France; Asturiana and Rubia Gallega cattle from Spain; and Italian Piedmontese cattle. They also analysed two Jersey cattle and two Holstein-Friesian cattle as 'controls'. Using a Polymerase Chain Reaction (PCR) and sequencing approach, they revealed seven DNA sequence variants within the coding region, including the previously described 11-nucleotide deletion and four DNA sequence variants in intron sequences that they claimed were 'probably neutral'. Dunner *et al.* (2003) studied 678 cattle from 28 European breeds using a Single Strand Conformational Polymorphism (SSCP) method, defining seven new sequence variations, and concluding that *MSTN* was highly variable having defined twenty haplotypes, with approximately one nucleotide substitution every 100 bp.

Ensemble (EMBL-EBI, Hinxton, United Kingdom (UK), Release 93 - July 2018) UMD3.1:GK000002.2 for cow *MSTN* lists 36 exon 1 variants, including missense and synonymous variations; 101 intron 1 variants, including variation in splice sites; 22 exon 2 variants; 86 intron 2 variants; and eight exon 3 variants. There is also extensive variation described in the promoter and upstream regions, 5' untranslated region, 3' untranslated region and further into the downstream regions.

Little effort has been made to characterise variation in *MSTN* in NZ cattle breeds, and while these breeds are predominantly derived from European/UK stock, little is known about how they relate to the current parent breeds. In this study, Polymerase Chain Reaction-Single Stranded Conformational Polymorphism (PCR-SSCP) analysis and DNA sequencing were used to investigate genetic variation in selected regions of *MSTN* in NZ cattle breeds.

2. Material and Methods

2.1. Cattle Investigated, DNA sample collection and purification

This research involving animals was performed in accordance with the Animal Welfare Act 1999 (NZ Government). A total of 722 *Bos taurus* cattle were investigated, with these being of a variety of breeds found in NZ including; Hereford (n = 30), Angus (n = 16), South

Devon (n = 5), Composite (n = 3), Charolais (n = 9), Red Poll (n = 19), Shorthorn (n = 18), Simmental (n = 15), Murray Grey (n = 17) and cross-bred Holstein-Friesian x Jersey cattle (n = 590). These animals did not have an observable double-muscled phenotype, but some of them were bred using eBVs to select phenotypically for improved growth and carcass traits. The cattle were genotyped in order to identify genetic variation across an extended region of *MSTN*. Of these ten breeds, only the HF x J-cross cattle were bred for milk production, with the other nine breeds being farmed primarily for meat production. The NZ Angus and Hereford are the preferred maternal breeds in NZ, partly because of their hardiness and suitability for farming in extensive hill and high country (rangeland) farming systems.

Blood samples were collected from each animal onto FTA cards (Whatman, Middlesex, UK) by piercing the ear of the animal. The samples were air dried and DNA extraction was carried out using a two-step procedure as described by Zhou *et al.* (2006).

2.2. Polymerase Chain Reaction (PCR)

The Polymerase Chain Reaction (PCR) was used to amplify five different regions of *MSTN* (Figure 1). The PCR was performed in 15 μ L reactions containing genomic DNA on a 1.2-mm diameter disc of FTA card, 0.25 μ M for each primer, 150 μ M for each dNTP (Eppendorf, Hamburg, Germany), 3.0 mM Mg^{2+} , 0.5 U of *Taq* DNA polymerase (Qiagen, Hilden, Germany), and the reaction buffer supplied with the enzyme.

Amplification was undertaken in Bio-Rad S1000 thermal cyclers (Bio-Rad, Hercules, CA, USA). The thermal cycling conditions included an initial denaturation at 94 °C for 2 min, followed by 34 cycles of 94 °C for 30s, annealing for 30s at 58 °C or 60 °C depending on primer pairs (see figure. 1), and a final extension step at 72 °C for 5 min.

The amplicons obtained from the PCR reactions were visualized by electrophoresis in 1% agarose (Quantum Scientific, Queensland, Australia) gels, using 1x TBE buffer (89 mM Tris, 89 mM boric acid, 2 mM Na_2EDTA), containing 200 ng/mL ethidium bromide. A 2 μ L aliquot of PCR product was added to 2 μ L of loading dye (0.2% bromophenol blue, 0.2% xylene cyanol, 40% (w/v) sucrose) and the gels were run at a constant 10 V/cm for 10 min., prior to visualization by UV transillumination at 254 nm.

2.3. Detection of genetic variation using Single Stranded Conformational Polymorphism (SSCP) analysis

The amplicons obtained from the PCR reactions were subjected to SSCP analysis. A 0.7 μ L aliquot of the amplicons was added to 7 μ L of loading dye containing 10 mM EDTA, 0.025% bromophenol blue, 0.025% xylene-cyanol, and 98% formamide. The samples were then denatured at 95 °C for 5 minutes, prior to snap chilling on wet ice. They were then loaded onto 16 cm x 18 cm, acrylamide: bisacrylamide (37.5:1) (Bio-Rad) gels. Electrophoresis was carried out using Protean II xi cells (Bio-Rad) for 19 hr at various voltages, gel concentrations and room temperatures as described in figure 1.

To detect the various homozygous and heterozygous SSCP patterns, the gels were silver stained using the method described by Byun *et al.* (2009).

2.4. Detection of genetic variation using nucleotide sequencing

Based on the PCR-SSCP patterns observed, cattle that were homozygous with unique banding patterns were subjected to direct sequencing. For heterozygous variants, the unique band was excised from the wet gel and incubated at 51 °C for 1 hr. A 1 μ L aliquot of the incubated product was aspirated and mixed with 14 μ L of PCR pre-mixture for amplification via PCR reaction. Using a MiniElute™ PCR Purification Kit (Qiagen), DNA from the PCR products were purified and sequenced by the Lincoln University DNA Sequencing Facility using the original PCR primers to prime the sequencing reactions. Each purified amplicon was sequenced in triplicate, in both directions, using the forward and reverse primers. The integrity of the sequence reactions were confirmed by viewing the peaks using Sequence Scanner v1.0 (Applied Biosystems). The sequence results were then subjected to subsequent sequence alignment and other comparisons using DNAMAN (Version 5.2.10, Lynnon BioSoft, Vaudreuil, Canada).

2.5. Determination of extended haplotypes

In determining the extended haplotypes of *MSTN*, two sets of primers were used to generate two amplicons (amplicons 2 and 4), and specific sequences were identified for each variant using the SSCP and sequencing methods described above. A combination of different PCR-SSCP pattern enabled the identification of the extended haplotypes using the following technique: cattle homozygous for one PCR-SSCP pattern in a region of the gene (e.g. *AA* for amplicon 2), were then genotyped in another region of the gene. If heterozygous in the second region (e.g. *AB* for amplicon 4), then two haplotypes (*A-A* and *A-B*), spanning the first and second regions could be defined.

3. Results

3.1. Sequence variation in bovine *MSTN*

In this study, five pairs of primers were used to genotype 722 cattle from ten different breeds of NZ cattle. A total of five amplicons covering 1,967 bp were generated. The amplified regions included exons 1, 2 and 3 and parts of introns 1 and 2 (Figure 1).

Five different PCR-SSCP analyses coupled with DNA sequencing were used to search for genetic variation in amplicons 1 - 5 in all the cattle investigated. This technique revealed seven banding patterns in amplicon 2 (intron 1 region) and three patterns in amplicon 4 (which encompasses the intron 2 - exon 3 boundary region). No genetic variation was observed in the other amplified regions. Of the seven variants identified in the intron 1 region, the Hereford and cross-bred Holstein-Friesian x Jersey cattle had five variants each (*A*, *B*, *C*, *D* and *E*) and (*A*, *B*, *C*, *D*, and *G*) respectively, while two variants (*A* and *B*) were observed in Simmental cattle, three variants (*A*, *B* and *C*) in Composite and Red Poll cattle, variants (*A*, *B* and *E*) in Angus and South Devon cattle, variants (*A*, *C* and *E*) in Charolais cattle, and four variants (*A*, *B*, *C* and *F*) and (*A*, *B*, *C* and *E*) in Shorthorn and Murray Grey cattle respectively. In the intron 2 - exon 3 region, all the breeds were observed to carry the *A*, *B* and *C* variants, except for Red Poll and Composite cattle, with only the *A* and *B* variants being detected.

A total of 13 nucleotide variations were identified across the regions examined. Twelve were nucleotide substitutions (c.373+751G>T, c.373+803T>G, c.373+877A>G, c.373+895G>C, c.374-909C>T, c.374-842G>C, c.374-812A>G located in intron 1, and c.748-195C>T, c.748-196C>T, c.748-281C>G, c.748-350C>T, and c.748-352C>T located in the intron 2 - exon 3 boundary region. One of the nucleotide variations was a deletion c.748-78del (Table 1). Of the twelve substitutions, five (c.373+751G>T, c.373+803T>G, c.373+877A>G, c.373+895G>C and c.748-350C>T) were novel and are reported here for the first time, whereas the remaining seven, with their respective rs numberings (c.374-909C>T

rs109597350, c.374-842G>C rs110496111, c.374+812A>G rs385362581, c.748-195C>T rs207514521, c.748-196C>T rs210984981, c.748-281C>G rs110910793, and c.748-352C>T rs132859209) had been previously reported to Ensembl (EMBL-EBI, Hinxton, United Kingdom, release 93 - July 2018 UMD3.1:GK000002.2 for cow *MSTN*). The deletion c.748-78del was identified at the intron 2 - exon 3 boundary and has been previously described by Grobet *et al.* (1998).

3.2. Extended haplotypes across regions of introns 1 and 2

A total of 18 extended haplotypes (H1-H18) were unambiguously defined spanning amplicon 2 (intron 1) and amplicon 4 (the intron 2 - exon 3 boundary region). Some of these haplotypes were exclusive to a breed, while others appeared to be common across the cattle breeds (see Table 2). The extended haplotype H11, carried c.373+803T>G, c.374-909C>T, c.374-842G>C and c.748-78del, and was only observed in the Hereford cattle. Similarly, extended haplotype H14 with c.373+803T>G, c.374-842G>C, c.748-350C>T, c.748-196C>T and c.748-195C>T, was observed only in the South Devon cattle. Haplotypes H12, H16, H17 and H18 were exclusive to the cross-bred Holstein-Friesian x Jersey cattle. Haplotype H2, bearing only the deletion c.748-78del was the most common, and was found in all breeds, except the South Devon cattle (Table 2).

4. Discussion

This study investigated *MSTN* in ten breeds of NZ *Bos taurus* cattle that are used for meat and/or milk production. A PCR-SSCP approach and DNA sequencing were used to characterise an extended region of the gene, spanning nucleotide c.-25 in the 5'UTR to c.*24 in the 3'UTR. This encompassed the three exons, and parts of introns 1 and 2. Using this approach we identified previously described sequence variation, as well as some novel variations.

Of the nucleotide variations identified, one was a single nucleotide deletion and twelve were substitutions. The deletion was located at position c.748-78del near the intron 2 - exon 3 boundary and is close to the splice-donor site. It has been previously identified in European cattle breeds (Grobet *et al.* 1998). It is becoming evident that introns are functionally active participants in gene and genome functionality. This is because they can encode regulatory elements (Yutzey *et al.* 1989) that participate in splicing, transcription, and recombination events. This means that both conserved and variable non-coding sequences in *MSTN* may

possess functionality. For example, in Latvian Darkhead sheep, Sjakste *et al.* (2011) identified sequence variation at the intron 1 boundary and suggested that variation can initiate single-strandedness and the formation of an additional hairpin bend in the intron. Such perturbation of the pre-mRNA secondary structure could potentially influence sequence interaction with different regulatory proteins and the efficiency of lariat formation, and thus potentially transcription and splicing efficiency (Chasin, 2007; Hiller *et al.* 2007; Aznarez *et al.* 2008; Wang and Burge, 2008).

Seven of the twelve nucleotide substitutions (c.374-909C>T; rs109597350, c.374-842G>C; rs110496111, c.374+812A>G; rs385362581, c.748-195C>T; rs207514521, c.748-196C>T; rs210984981, c.748-281C>G; rs110910793, and c.748-352C>T; rs132859209) have been documented by Ensembl, and two of these (c.748-281C>G and c.748-352C>T) have been reported previously (He *et al.* 2013). He *et al.* (2013) compared sequences from Qinchuan cattle with a *Bos taurus* sequence (GenBank accession number AF320998), reporting 69 sequence variations. Of these, one was a 16-bp insertion in intron 1, which has also been previously reported (Tantia *et al.*, 2006). The length, position and number of nucleotides in the intron 1 of *MSTN* in the Qinchuan cattle would increase as a result of this 16-bp insertion, and this possibly explains why none of the sequence variation identified in intron 1 of our study, was described in their report. He *et al.* (2013), also reported four sequence variations (c.373+807G>T, c.373+881A>G, c.373+899G>A and c.373+934C>T) that we did not find in this same region. Taken together, this would suggest that this region of *MSTN* is highly variable and possibly under less functional constraint, with this having also been alluded to in sheep (Hickford *et al.* 2009).

Since these variations in sheep and cattle occur in a non-expressed region of *MSTN*, it is perhaps less obvious how they might affect the structure or function of the protein. However, He *et al.* (2010) transformed C2C12 cell-lines with a transgene construct that contained part of the bovine *MSTN* promoter (pMD-MSTNPro) and another construct containing the first intron of bovine *MSTN* (pMD-Intron1), along with a reporter gene (Green Fluorescent Protein; GFP). They observed an increase in fluorescence and the number of fluorescence positive cells and concluded that the presence of intron 1 of bovine *MSTN* improved the expression of GFP in the transformed cells. Conversely, while there is little or no information on the possible effect of the absence of bovine *MSTN* intron 1 on gene expression, it would be interesting to explore this path, especially, since most of the previous studies (McPherron

et al., 1997) have only focused on examining the effects of “knocking-out” the active portion of the MSTN peptide, and not the introns.

There was no genetic variation identified in the coding sequences of bovine *MSTN* in this study. However, Dunner *et al.* (2003) identified five nucleotide variations in the coding region of bovine *MSTN*, two of which brought about an amino acid change, while the others were silent. These included p.S105C, which was described in Parthenaise cattle and p.D182N resulting from a G>A substitution in the Maine-Anjou breed. A silent exon 1 substitution c.267A>G was found in Aubrac, Bazadaise and Salers cattle, whereas c.324C>T in exon 1 was observed in Maine-Anjou, Charolais, Aubrac, and Salers cattle, and an Intr95 sire-line. A third silent substitution, c.387G>A, was found in exon 2 in Ayrshire, Maine-Anjou, Salers, and Galloway cattle.

A possible explanation for why exon variation was not found in the NZ cattle may be because of the breeds investigated. While Dunner *et al.* (2003), investigated a total of 28 breeds, only nine: the Parthenaise, Main-Anjou, Salers, Aubrac, Bazadaise, Galloway, INTR95 sire-line, Ayrshire and Charolais were reported to have variation in the coding regions of the gene. Of these nine breeds, only the Charolais was among the breeds investigated in our study, and the Charolais cattle investigated did not have the c.324C>T substitution. This could have been for a number of reasons including test sample size, founder effects, and selection including cross-breeding and in-breeding. Some of these reasons have been suggested previously by Blott *et al.* (1998), who revealed genetic differences between Hereford cattle from Britain, NZ, Canada, Ireland and Sweden. In a study by Kantanen *et al.* (2000b), it was suggested that genetic divergence in Nordic cattle breeds and within population diversity can be explained by the combined effects of breed origin, admixture during foundation and development, and random genetic drift due to limitations in effective population size, either when the breed was founded, or more recently. Therefore, while most of the NZ breeds might have originated from Europe, a number of different effects may have contributed to differences existing between these breeds and their European ancestors.

Some of the variants were found to be common to more than one breed, while others were apparently rare and breed-specific. For example, of the seven variants defined in intron 1, variant *A* was found in all the ten breeds and crosses. The Hereford cattle and the cross-bred Holstein-Friesian x Jersey cattle shared variants *A*, *B*, *C* and *D*, while the Composite, Red

Poll, Murray Grey and Shorthorn cattle shared the *A*, *B* and *C* variants. In the intron 2 - exon 3 region studied, the three variants were found in all the breeds except the Red Poll and Composite cattle, which carried only the *A* and *B* variants. Of the nine 'beef breeds' investigated, five (Hereford, Angus, South Devon, Murray Grey and Charolais) shared intron 1 *E*. These five breeds are among the most common beef breeds in NZ, hence sharing *E* may be indicative of among other things having a common ancestor, or from having been historically inter-bred, or because of selection for important meat/carcass-related traits. In the latter context, *E* might be associated with improved performance, although the absence of *E* in the other beef cattle (Composite cattle, Red Poll, Shorthorn and Simmental) might weaken this selection argument, and overall care needs to be taken in the context of only a small number of any of the breeds except the dairy cross-bred Holstein-Friesian x Jersey cattle, having been studied. Equally, these cattle were not, and probably could not have been selected so as to be representative of the whole breed population in NZ, as all the cattle studied were in private ownership.

While sequence variation may provide some indication of the structural diversity of a gene and uniqueness of a breed or breeds, extended haplotypes spanning a region are typically more informative, especially if they encompass all or most of the coding sequences. A total of 18 haplotypes (H1-H18) spanning intron 1 (amplicon 2) to the intron 2 - exon 3 boundary region (amplicon 4) of *MSTN* were resolved (Table 3). It has been shown that sequence variations that are adjacent or close to each other tend to be co-inherited and thus usually show strong linkage disequilibrium (Hey 2004). This phenomenon was observed with the *MSTN* haplotypes, as they could be separated into seven broad haplotype groups based on the location of the sequence variation. For example, with the intron 1 region, haplotypes H1-H6 carried c.374-909T (named Group 1), while H7-H9 with c.373+751T, c.373+803G c.373+877G, c.373+ 895C and c.374-909T cluster as group 2, whereas H10-H14 with nucleotide substitutions c.373+803G, c.374-909T and c.374-842C cluster as group 3, and H15-H18 with substitutions c.373+803G c.373+ 877G, c.374-909T and c.374-812G cluster as group 4. In the intron 2 region, haplotypes H3, H6, H9, H12, H14, H15 and H18 which carry c.748-195T, c.748-196T and c.748-350T could be clustered as group 5, H1, H4, H7, H10 and H16 could be clustered as group 6 with nucleotide substitutions/deletions c.748-78del, c.748-281G and c.748-352T, and finally H2, H5, H8, H11, H13 and H17 with c.748-78del could be clustered as group 7. Recombination activities do not occur randomly throughout the genome, but tend to occur in recombination hotspots ([Jeffreys et al. 2005](#)),

which are usually small regions characterized by significantly higher recombination rate than in surrounding regions. These hotspots are associated with trinucleotide repeats such as CGG-CCG, GAG-CTG, GAA-TTC, and GCN-NGC and are conserved in mammals and yeast (Aguilera *et al.* 2008). Investigations have suggested that these hotspots tend to form hairpin structures on the lagging strand during replication from single-stranded DNA base-pairing with itself in the trinucleotide repeat region (Aguilera *et al.* 2008). These hairpin structures cause DNA breaks that lead to a higher frequency of recombination at these sites (Aguilera *et al.* 2008). In our investigation, we observed some of these trinucleotides associated with DNA hotspots in the regions of introns 1 and 2 investigated. Some occurred as a single sequence, while others occurred as a repeat in the positions where our variations were identified. For example, the trinucleotide ‘CTG’ was found in the intron 1 position where c.373+751G>T was detected. We also observed the trinucleotide sequence ‘GAA-TTC’ in the region of intron-2 where the adjacent substitutions c.748-195C>T and c.748-196C>T were identified, but it was only the trinucleotide ‘TTC’ that seemingly appeared as a repeated sequence in the positions where substitutions c.748-350C>T, c.748-352C>T and c.748-78delT occurred. These trinucleotides were not observed as precisely repeated sequences in the regions where all the variations were found except for the ‘TTC’, hence it is difficult to conclude whether these regions of introns 1 and 2 are actual hotspots for DNA recombination. However, it is important to note that the identification of these trinucleotide sequence(s) in the exact positions where our proximal substitutions were detected may not have been by chance. This could suggest that these trinucleotide sequences are capable of influencing DNA recombination irrespective of whether they appear as a repeated sequence or as a single sequence. Further analysis of these regions of introns 1 and 2 will have to be undertaken to ascertain the possibilities of these regions serving as hotspots for DNA recombination.

The extended haplotype H2, carrying the deletion c.748-78del, occurred at a relatively high frequency in all the breeds except the South Devon cattle. The sharing of common haplotypes suggests that cattle might be more closely related, although this is no proof. Dunner *et al.* (2003) suggested that the pattern of haplotype sharing is an indicator of the history of the different bovine breeds, so the distribution of shared haplotypes is useful in investigating population relationships. The presence of breed-specific haplotypes on the other hand may explain the differences between breeds, especially if the haplotypes are found within the

coding sequences. In a study around the identification of shared haplotypes that are identical by descent (IBD), Ying *et al.* (2015), suggested that while common variants are mostly shared across ethnic groups, rare variants are more likely to be recent in history and population-specific. In this context, the occurrence of variants *F* and *G* is exclusively in the Shorthorn and cross-bred Holstein-Friesian x Jersey cattle respectively, suggests that these variants are breed-specific. Such variations occurring within a species might be expected, as they form the basis for individual differences within a population especially because, it is variation within both coding and non-coding sequences that produces phenotypic variation between both individuals in a species and between different species (Matic 2001). It will therefore, it will be interesting to see what unique trait these variants may be influencing in the respective breeds and whether their presence could explain the differences between these breeds.

The extended haplotype H15 found in the Shorthorn cattle, and H16, H17 and H18 observed in the cross-bred Holstein- Friesian and Jersey cattle (see Table 2), share some rare but also more common variations (Table 3). This is not surprising, especially considering the evolution and introduction of the Shorthorn breed into NZ, where it was initially used for milking purpose (Te Ara-the Encyclopaedia of NZ, accessed 16 November, 2018), just like the pure and cross bred Holstein Friesian and Jersey cattle of today. Since these breeds were farmed in NZ for the same purpose at some point in time, it is possible that, they may have shared a common ancestor, and this may explain why they share some *MSTN* variations in common.

In the region of intron 1 investigated, it is noteworthy that some breeds appear to have a reduced number of variants. For example, while the Hereford and the cross-bred Holstein-Friesian x Jersey cattle had five variants each, the Simmental had two variants, the Composite, Red Poll, Angus and South Devon had three variants each, and the Shorthorn and Murray Grey cattle were found to have four variants each. In the intron 2 - exon 3 region, all the breeds had three variants each, except for the Red Poll cattle and the Composite cattle, which had two variants. It is difficult to confirm whether this is a characteristic of the specific breeds, or a function of sampling and sample size. Interestingly, this phenomenon was also observed in sheep, as reported by Hickford *et al.* (2009) in a study involving 747 sheep in NZ, where only three of the five variants were observed in the NZ Romney sheep breed investigated. They suggested this could either be a unique characteristic of the breed, or

perhaps the sheep investigated may not have been a good representation of the breed across NZ, either by number or by the method of selection.

5. Conclusion

Based on the five novel substitutions identified in this study, together with the other seven substitutions and one deletion reported previously, it appears that *MSTN* is even more variable than reported previously. It might also therefore be expected that as more cattle breeds are investigated, more variation(s) unique to those breeds, or perhaps uncommon in other breeds, might be identified. Overall, this study provides more insight into the conserved and polymorphic nature of the coding and non-coding sequences of bovine *MSTN* respectively. It also highlights possible linkages of bovine *MSTN* haplotypes based on proximal occurrence of the observed sequence variation. Given the potential of this gene to affect muscle growth and thus meat production, the findings lay a foundation to proceed to see if this genetic variation might be of value in improving production traits.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Acknowledgement

The authors are very thankful to the academic and technical staff of Lincoln University for their support and facilities provided.

References

- Aguilera, A., and Gomez-Gonzalez, B., 2008. Genome instability: A mechanistic view of its causes and consequences. *Nat. Rev. Genet.* 9 (3), 204-17. DOI:10.1038/nrg2268.
- Aiello, D., Patel, K., and Lasagna, E., 2018. The myostatin gene: an overview of mechanisms of action and its relevance to livestock animals. *Anim. Genet.* 49, 505-519. DOI: 10.1111/age.12696
- Aznarez, I., Barash, Y., Shai, O., He, D., Zielenski, J., Tsui, L.C., Parkinson, J., Frey, B.J., Rommens, J.M., and Blencowe, B.J., 2008. A systematic analysis of intronic sequences downstream of 5' splice sites reveals a widespread role for U-rich motifs and TIA1/TIAL1 proteins in alternative splicing regulation. *Genome Res.* 18, 1247–1258.
- Barrett, L.W., Fletcher, S., and Wilton, S.D., 2012. Regulation of eukaryotic gene expression by the untranslated gene regions and other non-coding elements. *Cell. Mol. Life Sci.* 69 (21), 3613-3634.

- Bass, J.J., Baker, R.L., and Carter, A.H., 1976. Comparison of Angus, Charolais-Angus, South Devon-Angus, Hereford-Angus, and Friesian-Angus steers. *NZ. J. Exp. Agri.* 4 (1), 5-8, DOI: 10.1080/03015521.1976.10425836.
- Blott, S., Williams, J., and Haley, C., 1998. Genetic variation within the Hereford breed of cattle. *Anim. Genet.* 29, 202-211. DOI: 10.1046/j.1365-2052.1998.00326.x.
- Byun, S., Fang, Q., Zhou, H., and Hickford, J.G.H., 2009. An effective method for silver-staining DNA in large numbers of polyacrylamide gels. *Anal. Biochem.* 385(1), 174-175.
- Chasin, L., 2007. Searching for splicing motifs. *Adv. Exp. Med. Biol.* 623, 85-106.
- Dunner, S., Miranda, M.E., Amigues, Y., Canon, J., Georges, M., Hanset, R., Williams, J. and Menissier, F., 2003. Haplotype diversity of the myostatin gene among beef cattle breeds. *Genet. Sel. Evol.* 35, 103-118.
- Ensemble (EMBL-EBI, Hinxton, United Kingdom (UK). Released 93 - July 2018) UMD3.1:GK000002.2 for cow MSTN (<https://asia.ensembl.org/index.html>).
- Grobet L., Royo Martin L.J., Poncelet D., Pirottin D., Brouwers B., Riquet J., Schoeberlein A., Dunner S., Menissier F., Massabanda J., Fries R., Hanset R., Georges M.A., 1997. Deletion in the myostatin gene causes double muscled phenotype in cattle. *Nat. Genet.* 17, 71-74.
- Grobet, L., Poncelet, D., Royo, L.J., Brouwers, B., Pirottin, D., Michaux, C., Me'nissier, F., Zanotti, M., Dunner, S., and Georges, M., 1998. Molecular definition of an allelic series of mutations disrupting the myostatin function and causing double-muscling in cattle. *Mamm. Genome* 9, 210-213.
- Grisolia, A.B., Angelo, G.T.D., Porto-Neto, L.R., Siqueira, F., and Garcia, J.F., 2009. Myostatin (GDF8) single nucleotide polymorphisms in Nellore cattle. *Genet. Mol. Res.* 8, 822-830.
- González-Cadavid N., Taylor W., Yarasheski K., Sinha-hikim I., Ma K., Ezzat S., Shen R., Lalani R., Asa S., Mamita M., Nair G., Arver S., Bhasin S., 1998. Organization of the human myostatin gene and expression in healthy men and HIV-infected men with muscle wasting. *Proc. Natl. Acad. Sci. USA*; 95, 14938-14943.
- Han, J., Forrest, R.H., Sedcole, J.R., and Hickford, J.G.H., 2015. Myostatin (MSTN) gene haplotypes and their association with growth and carcass traits in New Zealand Romney lambs. *Small Rum. Res.* 127, 8-19. DOI: <http://dx.doi.org/10.1016/j.smallrumres.2015.03.015>
- Han, J., Zhou, H., Forrest, R.H., Sedcole, J.R., Frampton, C.M., and Hickford, J.G.H., 2010. Effect of Myostatin (*MSTN*) g+6223G>A on production and carcass traits in New Zealand Romney sheep. *Asian-Aust. J. Anim. Sci.* 23, 863-866.
- He, Y.L., Wu, Y.H., Quan, F.S., Liu, Y.G., and Zhang, Y., 2010. Molecular analysis of the first intron in the bovine myostatin gene. *Mol. Biol. Rep.* 38, 4643-4649.
- He, Y.L., Wu, Y.H., Quan, F.S., Liu, Y.G., and Zhang, Y., 2013. Comparative analysis of myostatin gene and promoter sequences of Qinchuan and Red Angus cattle. *Genet. Mol. Res.* 12 (3), 3398-3406.

- Hey, J., 2004. What's So Hot about Recombination Hotspots? PLoS Biol. 2(6): e190. <https://doi.org/10.1371/journal.pbio.0020190>.
- Hickford, J.G.H., Forrest, R.H., and Zhou, H., 2009. Association between a g. +6723G-A SNP in the myostatin gene (*MSTN*) and carcass traits in New Zealand Texel sheep. J. Anim. Sci. 87, 1853.
- Hickford, J.G.H., Forrest, R.H., Zhou, H., Fang, Q., Han, J., Frampton, C.M., and Horrell, A.L., 2010. Polymorphisms in the ovine myostatin gene (*MSTN*) and their association with growth and carcass traits in New Zealand Romney sheep. Anim. Genet. 41, 64-72.
- Hiller, M., Zhang, Z., Backkofen, R., and Stamm, S., 2007. Pre-mRNA secondary structures influence exon recognition. PLoS Genet. 3, e204
- Jeffreys, A.J., Neumann, R., Panayi, M., Myers, S., Donnelly, P., 2005. Human recombination hot spots hidden in regions of strong marker association. Nat. Genet. 37, 601-606.
- Kambadur, R., Sharma, M., Smith, T.P., and Bass, J.J., 1997. Mutations in myostatin (GDF8) in double-muscle Belgian Blue and Piedmontese cattle. Genome Res. 7, 910-916.
- Kantanen, J., Olsaker, I., Holm, L.E., Lien, S., Vilkkilä, J., Brusgaard, K., Eythorsdottir, E., Danell, B., Adalsteinsson, S., 2000b. Genetic Diversity and Population Structure of 20 North European Cattle Breeds. J. Hered. 91, 446-457.
- Mattick, J.S., 2001. Non-coding RNAs: The architects of eukaryotic complexity. EMBO Rep. 2, 986-991.
- McPherron, A.C., and Lee, S.J., 1997. Double muscling in cattle due to mutations in the myostatin gene. Proc. Natl. Acad. Sci. 94, 12457-12461.
- Robert Peden, 'Beef farming', Te Ara - the Encyclopedia of New Zealand, <http://www.TeAra.govt.nz/en/beef-farming/print> (accessed 16 November 2018).
- Smith, J.A., Lewis, A.M., Wiener, P., and Williams, J.L., 2000. Genetic variation in the bovine myostatin gene in UK beef cattle: allele frequencies and haplotype analysis in the South Devon. Anim. Genet. 31, 306-309.
- Sjakste, T., Paramonova, N., Grislis, Z., Trapina, I., and Kairisa, D., 2011. Analysis of the single-nucleotide polymorphism in the 5'UTR and part of intron 1 of the sheep myostatin gene. DNA and Cell Biol. 30, 433-444.
- Tantia, M.S., Vijh, R.K., Kumar, S.T., Mishra, B., and Reecy, J.M., 2006. Comparative analysis of GDF 8 (myostatin) in *Bos indicus* and *Bos taurus*. DNA Seq. 17, 311-313.
- Watterson, G.A., Guess, H.A., 1977. Is the most frequent allele the oldest? Theor. Popul. Biol. 11, 141-160.
- Wang, Z., and Burge, C.B., 2008. Splicing regulation: from a parts list of regulatory elements to an integrated splicing code. RNA 14, 802-813.

Ying, D., Sham, P.C., Smith, D.K., Zhang, L., Lau, Y.L., Yang, W., 2015. HaploShare: Identification of extended haplotypes shared by cases and evaluation against controls. *Genome Biol. (Online Edition)*, 16, (92). <http://hdl.handle.net/10722/211876>

Yutzey, K.E., Kline, R.L., and Konieczny, S.F., 1989. An internal regulatory element controls troponin I gene expression. *Mol. Cell Biol.* 9, 1397–1405.

Zhou, H., Hickford, J.G.H., and Fang, Q., 2006. A two-step procedure for extracting genomic DNA from dried blood spots on filter paper for polymerase chain reaction amplification. *Anal. Biochem.* 354, 159-161.

Zhou, H., Hickford, J.G.H., and Fang, Q., 2008. Variation in the coding region of the myostatin (GDF8) gene in sheep. *Mol. Cellular Probes.* 22, 67-68.